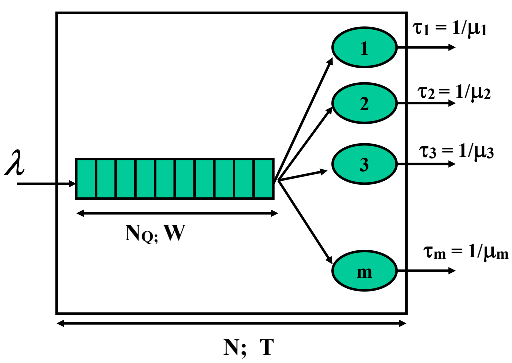
Materials covered

- M/M/1 Example
- Queuing Models
- M/M/∞ (infinite servers)
- M/M/m/m (finite servers and finite buffer size)
- M/M/m (m servers; infinite buffer size)
- Examples

Queuing Models

- Useful in performance evaluation of networks
- Queuing Process Notation: A/B/C/D
  - A: Arrival Process distribution: M for
  - B: Service Process Distribution: Common Exponential
  - C: Number of servers
  - D: Customer size allowed in the system
- System Parameter Notations:
  - N: Average number of customers in the system
  - T: Average customer time in the system
  - N_Q: Average number of customers waiting in queue
  - W: Average customer waiting time in queue

Model To be studied



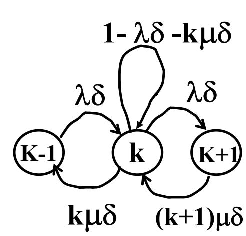
Model Assumptions for M/M/m Q

- Arrival process $\{A(t)|t>0\}$: Poisson with rate λ
- Service process:
  - Server i has service rate μ_i (for identical servers $\mu_i = \mu$)
  - System service rate depends on how many servers are busy
  - Assume k servers are busy
  - The next departure time is given by
    $X = \min(\tau_1, \tau_2, \ldots, \tau_k)$
    $P[X>t] = P[\min(\tau_1, \tau_2, .., \tau_k) > t]$
    $= \Pi\, P[\tau_i > t] = \Pi\, e^{-\mu_i t}$ (for identical servers $= e^{-k\mu t}$)
  - If **k identical** servers are busy, the service rate is $k\mu$

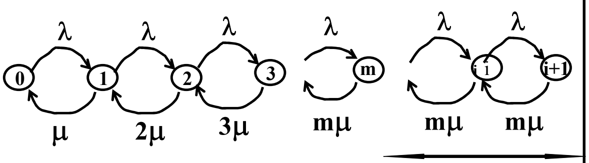
Arrival-Service Process of M/M/k Queue

- Consider a negligibly small interval $(0, \delta]$
- Joint service time τ of k servers is exponential with parameter $k\mu$
- $P[\tau \le \delta] = 1 - e^{-k\mu\delta} = (1 - (1 - k\mu\delta + (k\mu\delta)^2/2! + ..)) = k\mu\delta + O(\delta)$
- P[0 arrival, 0 departure]$=P[A(\delta)=0, \tau \ge \delta]$
  $=P[A(\delta)=0]P[\tau \ge \delta] = (1-\lambda\delta) * (1-k\mu\delta) = 1 - \lambda\delta - k\mu\delta + O(\delta)$
- P[0 arrival, 1 departure]$= P[A(\delta)=1, \tau \le \delta] = (1-\lambda\delta)\, k\mu\delta$
  $= k\mu\delta + O(\delta)$
- P[1 arrival, 0 departure]$= P[A(\delta)=1, \tau \ge \delta] = \lambda\delta * (1-k\mu\delta)$
  $= \lambda\delta + O(\delta)$
- Also note that
  - P[1 arrival, 1 departure]$= P[A(\delta)=1, \tau \le \delta] = \lambda\delta * \mu\delta = O(\delta)$
  - P[$\ge$ 1 arrival, $\ge$ 1 departure]$= O(\delta)$ P[$\ge$1 departure]$= O(\delta)$

Probability Transitions of State <u>k</u> in time interval span of <u>δ</u>

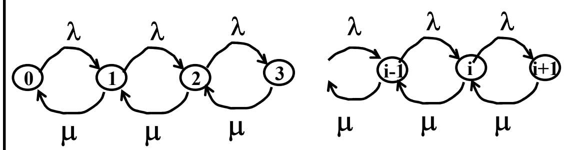


State Diagram of the M/M/m Queuing system



Service rate is $m\mu$
When system has m or more customers

Modeling of M/M/1Queue



State Diagram of the M/M/1 Queuing system

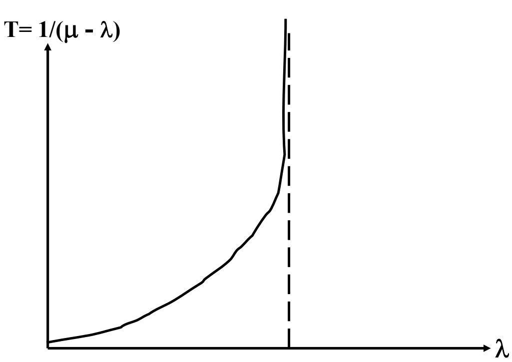
Derivation of Steady State Q Probabilities

- Let **n** = number of customers in the system
- Let p_n = steady state probability of n customers in system
- Writing Global Balance equation for **n=0** leads to
  $p_0 \lambda = p_1 \mu \Rightarrow p_1 = \rho\, p_0$ (here $\rho = \lambda / \mu$)
- Writing Global Balance equation for **n=1** leads to
  $p_0 \lambda + p_2 \mu = p_1 (\lambda + \mu) \Rightarrow p_2 \mu = p_1 \lambda \Rightarrow p_2 = \rho\, p_1$
- Writing Global Balance equation for **n=i** leads to
  $p_{i-1} \lambda + p_{i+1} \mu = p_i (\lambda + \mu) \Rightarrow p_i \mu = p_{i+1} \lambda \Rightarrow p_{i+1} = \rho\, p_i$
- Note ($p_0 + p_1 + p_2 + \ldots + \infty$ terms) = 1
  $\Rightarrow p_0(1 + \rho + \rho^2 + \rho^3 + ..) = 1 \quad \Rightarrow \rho < 1 \Rightarrow \lambda < \mu$
- If $\lambda < \mu$ **then** $p_0/(1-\rho) = 1$
  $\Rightarrow p_0 = (1-\rho) \qquad \Rightarrow p_i = (1-\rho)\,\rho^i$

Derivation Q Parameters

- Average number of customers in the system
  $N = \Sigma n p_n = \Sigma\, i\, \rho^i (1-\rho) = (1-\rho)\, \rho\, d/d\rho\{1/(1-\rho)\} = \rho/(1-\rho) = \lambda/(\mu - \lambda)$
- **Using Little's theorem** $N = \lambda T$ **leads to** $T = 1/(\mu - \lambda)$
- $T = (W + 1/\mu)$ **leads to** $W = \lambda/\mu(\mu - \lambda) = \rho/(\mu - \lambda)$
- Number of customers in Queue $N_Q = \lambda W = \rho^2/(\mu - \lambda)$
- **Comments**
  - **Note that** $p_0 = (1-\rho)$ is the proportion of the time system is idle
  - **Think of** ρ **as** the fraction of the time system is utilized: **utilization factor**

Behavior of delay

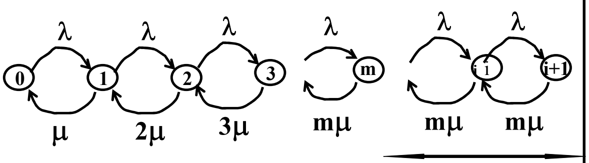


Example of M/M/1

- Comparison of server powers
  - **Q_1 has parameters (λ,μ)**
  - **Q_2 has parameters $(K\lambda,K\mu)$ with $K > 1$**
  - **From the table below:**
    - **Average number of customers waiting in queue is same but the Q2 is faster by a factor of K**

Q	ρ	N	Delay T
Q_1	λ/μ	$\rho/(1-\rho)$	$1/(\mu - \lambda)$
Q_2	λ/μ	$\rho/(1-\rho)$	$1/K(\mu - \lambda)$

Modeling M/M/m Queuing system



Service rate is $m\mu$
When system has m or more customers

Derivation of Steady State Q Probabilities

- We can view this as a truncated M/M/m/m Q with variable service rate concatenated with a M/M/1 Q with service rate **$m\mu$.**
- With previous notations, Global Balance equation for **n=0: $p_0\,\lambda = p_1\,\mu \Rightarrow p_1 = m\rho\, p_0$** (here $\rho = \lambda / m\mu$)
- For **n=1: $p_0\,\lambda + p_2\, 2\mu = p_1\,(\lambda+\mu) \Rightarrow p_2(2\mu) = p_1\,\lambda$**
  $\Rightarrow p_2 = (m\rho)^2\, p_0/2!$
- Writing Global Balance equation for **n=i** leads to
  $p_i = p_0\,(m\rho)^i/i!\quad (i \le m)$
  $p_i = p_0 m^i \rho^i/m!\quad (i \ge m)$

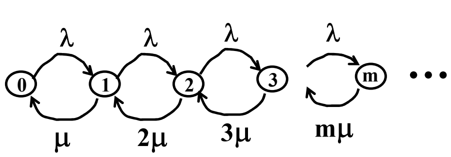
M/M/m Q model

- **Note ($p_0 + p_1 + p_2 + \ldots + \infty$ terms) = 1**
  $\Rightarrow p_0\{1+\ \Sigma\,(m\rho)^i/i!\ +\Sigma\,(m\rho)^i/i!\} = 1$
  $\Rightarrow p_0 = \{\Sigma\,(m\rho)^i/i!\ + (m\rho)^m/(m!(1-\rho)\}$
- **The probability that arriving customer will be queued is {Erlang's C formula}**
  $P\{N(t) \ge m\} = \Sigma\, p_i = \Sigma\, p_0 m^m \rho^i/i!$
  $P_Q = p_0(m\,\rho)^m/m!\ \Sigma\rho^{i-m} = p_0(m\,\rho)^m/\{m!(1-\rho)\}$
- **Average number of customers in the queue alone is**
  $N_Q = \Sigma i p_{m+i} = p_0(m\rho)^m/m!\ \Sigma i\rho^i = P_Q\,\rho/(1-\rho)$
- **Note $N_Q/P_Q = \rho/(1-\rho)$ As long as the arriving customer has to be in the queue, the M/M/m queue behaves as M/M/1 queue**

M/M/m Q model

- **Wait queue size is**
  - **$W = N_Q/\lambda = P_Q\,\rho/\lambda(1-\rho)$**
- **Average customer delay**
  - **$T = 1/\mu + W = 1/\mu\ + P_Q\,\rho/\lambda(1-\rho)$**
- **Average number of customers in the system**
  - **$N = \lambda T = m\rho + P_Q\,\rho/(1-\rho)$**

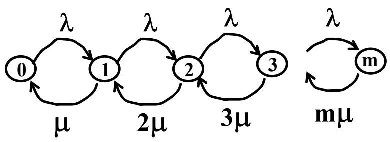
State Diagram of the $M/M/\infty$ Queuing system



$M/M/\infty$ Q model

- **Consider the result of M/M/m system**
  - **Global balance equation for N(t)=i is**
    - **$p_{i-1}\lambda = (i)\mu\, p_i\quad i=0,1,\ldots$**
    - **$p_{i-1} = p_0(\lambda/\mu)^i/i!\quad i=0,1,\ldots$**
- **Note this is Poisson Process!**
  $\Rightarrow p_0(1+\rho+\rho^2+\rho^3+..) = 1 \Rightarrow p_0 = e^{-\lambda/\mu} = e^{-\rho}$
  $\Rightarrow p_i = \rho^i\, e^{-\rho}$
- **Average Customers in the system**
  $N = \rho$ {Average of the Poisson Process}
- **System delay $T = N/\lambda = 1/\mu$**

Truncated Queue: M/M/m/m



State Diagram of the M/M/m/m Queuing system

M/M/m/m Q model

- **Global balance equation for N(t)=i is**
  - $p_{i-1}\lambda = (i)\mu\, p_i$ i=0,1,…m
  - $p_i = p_0(\lambda/\mu)^i/i!$ i=0,1,…m
- **Note this is Poisson Process!**
  - ➔ $p_0(1 + \rho + \rho^2/2! + \rho^3/3! + \ldots + \rho^m/m!) = 1$
  - ➔ $p_0 = [\Sigma\rho^i/i!]^{-1}$
  - ➔ $p_h = (\lambda/\mu)^h/h!\ [\Sigma\rho^i/i!]^{-1}$ h=0,1,…m
  - ➔ $p_m = (\lambda/\mu)^m/m!\ [\Sigma\rho^i/i!]^{-1}$ {Erlang's Blocking Formula}